Μια συχνή ανάθεση τιμής για το p είναι απλά ο αριθμός των διαφορετικών διακριτών τιμών ενός χαρακτηριστικού, άρα στο προηγούμενο παράδειγμα όπου έχουμε 3 τιμές (Ηλιόλουστος, Συννεφιασμένος, Βροχερός) έχουμε p = 1/3. Το m δρα σαν να καθορίζει πόσες επιπλέον παρατηρήσεις της συγκεκριμένης τιμής με πιθανότητα p προσθέτουμε στο δείγμα.

Βάσει των παραπάνω, στο προηγούμενο παράδειγμα έχουμε:

- Κλασσικός τρόπος: n_c=0, n=5 και συνεπώς P(*Συννεφιασμένος* | *Όχι*) = 0/5 =0.
- m-Εκτίμηση: n_c=0, n=5, p=1/3, m=3,

και συνεπώς P(*Συννεφιασμένος* | *Όχι*) = (0+3·(1/3))/5+3 = 1/8 = 0.125.

18.9 Μηχανές Διανυσμάτων Υποστήριξης

Οι *μηχανές διανυσμάτων υποστήριξης* ή *ΜΔΥ* (*Support Vector Machines*, *SVM*) προτάθηκαν από τον Vladimir Vapnik ως γραμμικοί ταξινομητές, το 1963. Απέκτησαν δημοσιότητα μετά το 1992, όταν ενισχύθηκαν από τον ίδιο και τους συνεργάτες του με το *κόλπο του πυρήνα* (*kernel trick*), που επέτρεψε τη χρήση τους και σε μη γραμμικώς διαχωρίσιμα προβλήματα. Οι SVM βασίζονται στη *θεωρία στατιστικής μάθησης* (*statistical learning theory*).

Στις δεκαετίες πριν και μετά το 2000, οι SVM εδραιώθηκαν ως μια από τις πιο διαδεδομένες μεθόδους (γραμμικής και μη) ταξινόμησης και παρεμβολής. Μέχρι την έλευση των βαθέων νευρωνικών δικτύων αποτελούσαν συνήθως τη βέλτιστη επιλογή για εφαρμογές όπως η αναγνώριση γραφής (handwriting recognition), η ταξινόμηση κειμένων (text categorization), η ταξινόμηση δεδομένων έκφρασης γονιδίων (gene expression data) και γενικά για εφαρμογές με δεδομένα πολλών διαστάσεων.

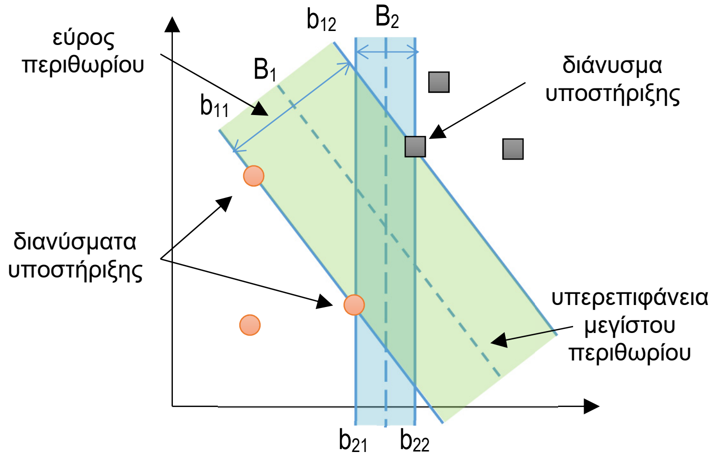
18.9.1 Η Βασική Ιδέα των SVM

Θεωρώντας ένα πρόβλημα δυαδικής ταξινόμησης με θετικά και αρνητικά παραδείγματα, μια επιφάνεια που λειτουργεί ως σύνορο μεταξύ των κλάσεων χαρακτηρίζεται από το γεγονός ότι χωρίζει τα θετικά από τα αρνητικά παραδείγματα. Τέτοια σύνορα υπάρχουν εν γένει πολλά, όπως για παράδειγμα τα B_1 και B_2 στο Σχήμα 18.16.

Η μέθοδος SVM επιδιώκει να βρει το σύνορο που απέχει όσο το δυνατόν περισσότερο από τα παραδείγματα των κλάσεων που διαχωρίζει. Η υπερεπιφάνεια αυτή ονομάζεται *υπερεπιφάνεια μεγίστου περιθωρίου* (*maximum margin hypersurface*) και σε γραμμικώς διαχωρίσιμα προβλήματα ορίζεται από έναν πεπερασμένο αριθμό παραδειγμάτων του συνόλου εκπαίδευσης που ονομάζονται *διανύσματα υποστήριξης* (*support vectors*). Με την παραπάνω προσέγγιση, στο πρόβλημα στο Σχήμα 18.16, το σύνορο B_1 είναι το σύνορο με το μέγιστο εύρος περιθωρίου.

Επιπλέον, μέσω των *συναρτήσεων πυρήνα* (*kernel functions*), οι SVM μπορούν να μετασχηματίσουν τον αρχικό χώρο υποθέσεων έτσι ώστε μη-γραμμικώς διαχωρίσιμα προβλήματα να μετατραπούν σε γραμμικώς διαχωρίσιμα και τελικά να λυθούν με την ίδια μεθοδολογία.

Επομένως, το κύριο πλεονέκτημα των SVM είναι ότι όχι μόνο δημιουργούν ένα διαχωριστικό σύνορο μεταξύ των επιμέρους κλάσεων (όπως άλλωστε κάνουν οι περισσότερες μέθοδοι ταξινόμησης), αλλά φροντίζουν αυτό το σύνορο να απέχει όσο το δυνατό περισσότερο από τα παραδείγματα των κλάσεων που διαχωρίζουν.



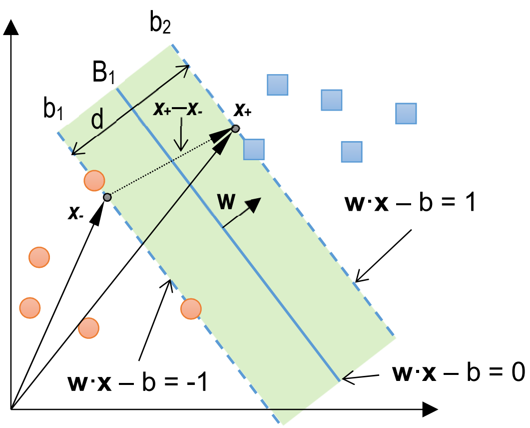
Σχήμα 18.16: Δύο σύνορα απόφασης (διακεκομμένες γραμμές) σε πρόβλημα δυαδικής ταξινόμησης. Το σύνορο B_1 χωρίζει τέλεια τις δύο κλάσεις και εξασφαλίζει επιπλέον και μέγιστο εύρος περιθωρίου.

Τα σύνορα απόφασης με μεγάλα περιθώρια έχουν κατά βάση μεγαλύτερη ανοχή σε φαινόμενα υπερπροσαρμογής. Αν το περιθώριο είναι μικρό, τότε μια μικρή διαταραχή (π.χ. εξαιτίας θορύβου στα δεδομένα εκπαίδευσης) μπορεί να προκαλέσει σημαντική πτώση στην απόδοση του ταξινομητή. Στα μεγάλα περιθώρια, ακόμη κι αν τα διανύσματα υποστήριξης είναι παραδείγματα με λίγο θόρυβο, τα σφάλματα γενίκευσης είναι κατά βάση μικρότερα.

18.9.2 Γραμμικές SVM

Γραμμικώς Διαχωρίσιμα Προβλήματα

Θεωρούμε ένα σύνολο από πολυδιάστατα σημεία $\boldsymbol{x}$ (x_1, x_2, ..., x_n) για τα οποία γνωρίζουμε ότι ανήκουν στην κλάση (+) (τετράγωνα) ή στην κλάση (–) (κύκλοι), τις οποίες έστω ότι τις κωδικοποιούμε αριθμητικά με 1 και –1 αντίστοιχα.



Σχήμα 18.17: Όρια απόφασης και περιθωρίου σε γραμμική SVM.

Θεωρώντας, για εποπτικούς λόγους, ένα γραμμικώς διαχωρίσιμο πρόβλημα σε δύο διαστάσεις (Σχήμα 18.17), η εξίσωση του συνόρου θα έχει τη μορφή $w_1x_1 + w_2x_2+b=0$ ή $\boldsymbol{w}\cdot\boldsymbol{x}+b=0$ αν υιοθετήσουμε διανυσματική γραφή για τα βάρη και το σημείο. Η τελευταία σχέση είναι αρκετά βολική καθώς αποκρύπτει την απλοποίηση των δύο διαστάσεων που υιοθετήσαμε στο Σχήμα 18.17. Εφαρμόζοντας τη σχέση αυτή για δύο σημεία $\boldsymbol{\alpha}$ και $\boldsymbol{\beta}$ που βρίσκονται πάνω στο σύνορο και αφαιρώντας τις δύο σχέσεις προκύπτει:

$$\boldsymbol{w}\cdot(\boldsymbol{x}_\alpha-\boldsymbol{x}_\beta)=0$$

που σημαίνει ότι το διάνυσμα των βαρών είναι κάθετο στο ζητούμενο σύνορο.

Σε γραμμικώς διαχωρίσιμα προβλήματα, μπορούμε να ορίσουμε δύο επιπλέον σύνορα εκατέρωθεν του $\boldsymbol{w}\cdot\boldsymbol{x}+b=0$ που επίσης χωρίζουν τις δύο κλάσεις και απέχουν όσο γίνεται περισσότερο μεταξύ τους. Η περιοχή που οριοθετούν αυτές οι δύο ευθείες (υπερεπιφάνειες, στη γενική περίπτωση) είναι το ζητούμενο μέγιστο περιθώριο (margin) και η ευθεία $\boldsymbol{w}\cdot\boldsymbol{x}+b=0$ (υπερεπιφάνεια) βρίσκεται στη μέση αυτού του περιθωρίου. Επιπλέον, για ένα σημείο $\boldsymbol{z}$ στο χώρο του προβλήματος, η κλάση y του σημείου θα είναι:

$$y = \begin{cases} 1 & \alpha\nu\ \boldsymbol{w}\cdot\boldsymbol{z}+b>0 \\ -1 & \alpha\nu\ \boldsymbol{w}\cdot\boldsymbol{z}+b<0 \end{cases}$$

Έστω $\boldsymbol{x}_+$ και $\boldsymbol{x}_-$ είναι σημεία στα πάνω και κάτω όρια του περιθωρίου αντίστοιχα (Σχήμα 18.17). Καθώς το διάνυσμα $\boldsymbol{w}$ είναι κάθετο στο ζητούμενο σύνορο, διαιρώντας το με το μήκος του $||\mathbf{w}||$ προκύπτει ένα μοναδιαίο διάνυσμα που πολλαπλασιαζόμενο (εσωτερικό γινόμενο) με το διάνυσμα της διαφοράς ($\boldsymbol{x}_+ - \boldsymbol{x}_-$), δίνει το ζητούμενο d:

$$(\boldsymbol{x}_+ - \boldsymbol{x}_-)\cdot\frac{\boldsymbol{w}}{||\boldsymbol{w}||} = d$$

Σε αυτή τη σχέση, για το $\boldsymbol{x}_+\cdot\boldsymbol{w}$, η σχέση υπολογισμού του y δίνει $1-b$ ενώ για το $\boldsymbol{x}_-\cdot\boldsymbol{w}$ δίνει $1+b$. Αντικαθιστώντας στην προηγούμενη σχέση προκύπτει:

$$d = \frac{2}{||\boldsymbol{w}||}$$

Άρα η μεγιστοποίηση του d ισούται με την ελαχιστοποίηση του $||\mathbf{w}||$.

Άρα η διαδικασία εκπαίδευσης της SVM μπορεί να οριστεί μαθηματικά ως η μεγιστοποίηση του d και ισούται με την ελαχιστοποίηση του $||\mathbf{w}||$ που πιο βολικά μπορεί να γραφεί ως:

$$\min_{\boldsymbol{w}}\frac{||\boldsymbol{w}||^2}{2}$$

Η τελευταία σχέση μαζί με τους περιορισμούς που ορίζει η έκφραση για τα y, αποτελεί τη μαθηματική διατύπωση της διαδικασίας εκπαίδευσης της SVM. Η επίλυση αυτού του προβλήματος γίνεται με τη βοήθεια των πολλαπλασιαστών Lagrange (*Lagrange multipliers*) και προκύπτει ότι τα $\mathbf{w}$ και b που ορίζουν το σύνορο απόφασης εξαρτώνται μόνο από διανύσματα (σημεία) που βρίσκονται πάνω στις ευθείες b_1 και b_2, και από εσωτερικά γινόμενα με αυτά. Τα διανύσματα αυτά ονομάζονται *διανύσματα υποστήριξης* (*support vectors*).

Μη γραμμικώς διαχωρίσιμα προβλήματα

Η κλασσική μέθοδος SVM αναζητά εκείνο το περιθώριο που διαχωρίζει τέλεια όλα τα δεδομένα της μιας κλάσης από τα δεδομένα της άλλης. Σε προβλήματα όμως του πραγ-